MS211, Cálculo Numérico, Turmas L, M
Segundo Semestre de 2020, UNICAMP

Resoluçao Atividade 01

Entrega dos exercícios SOMENTE por Google Classroom até terça 29/09/2020. Os exercícios podem ser desenvolvidos em grupos de até 3 (três) membros. Escreva o nome e o RA dos membros do grupo em todas as folhas, com destaque na primeira página. É aconselhável que somente um membro por grupo faça a entrega da atividade completa do grupo no Google Classroom.

Note que usamos a notação FP(β, p, l, u) para a aritmética finita em base β com p dígitos na mantissa e com o expoente que está no intervalo $[l, u]$.

(1)
a) Considere o sistema de ponto flutuante FP$(10, 4, -9, 9)$ que não use nem o *truncamento* nem o *arredondamento*. Quais dos seguintes números têm representação exata neste sistema?
10, -101, 0.000125, 0, 24.57400, -2220.2, 0.0013254, $-1.5 * 10^{-11}$, $-1.5 * 10^{10}$, $1.5 * 10^{-1}$, 45564.9, -0.000001, -0.00000000011, 998999999 e 99969999
Qual será a representação desses números no sistema FP quando usamos o *truncamento* e quando usamos o *arredondamento*?

b) Considere os três números $x = 9,274 * 10^5$; $y = 23,2$; $z = -2,3456 * 10^{-2}$. Compute na aritmética finita FP$(10, 3, -9, 9)$ que usa o arredondamento, a operação $x * z - 2 * (x + y)$. Que valor toma este calculo? Escreva cada passagem que usou para obter o resultado. Qual é o erro absoluto e o erro relativo com respeito à computação feita com a sua calculadora?

Resoluçao (1):

a)

Número	Trunc.	Arred	Repr. Exata?
10	$0.1 * 10^2$	$0.1 * 10^2$	Sim
-101	$-0.101 * 10^3$	$-0.101 * 10^3$	Sim
0.000125	$0.125 * 10^{-3}$	$0.125 * 10^{-3}$	Sim
0	Num. especial 0 or presisão maquina	*	Não/Sim
24.57400	$0.2457 * 10^2$	$0.2457 * 10^2$	Não
-2220.2	$-0.222 * 10^4$	$-0.222 * 10^4$	Não
0.0013254	$0.1325 * 10^{-2}$	$0.1325 * 10^{-2}$	Não
$-1.5 * 10^{-11}$	$-0.15 * 10^{-10}$(*não representavel, underflow*)	*underflow*	Não
$-1.5 * 10^{10}$	$-0.15 * 10^{11}$(*não representavel, overflow*)	*overflow*	Não
$1.5 * 10^{-1}$	$0.15 * 10^0$	$0.15 * 10^0$	Sim
45564.9	$0.4556 * 10^5$	$0.4556 * 10^5$	Não
-0.000001	$-0.1 * 10^{-5}$	$-0.1 * 10^{-5}$	Sim
-0.00000000011	$-0.11 * 10^{-9}$	$-0.11 * 10^{-9}$	Sim
998999999	$0.9989 * 10^9$	$0.9990 * 10^9$	Não
99969999	$0.9996 * 10^8$	$0.9997 * 10^8$	Não

b) $x = 9,274 * 10^5$; $y = 23,2$; $z = -2,3456 * 10^{-2}$
Na aritmética finita $FP(10,3,-9,9)$ com arreondamento:
$\bar{x} = 0.927 \cdot 10^6$, $\bar{y} = 0.232 \cdot 10^2$, $\bar{z} = -0.235 * 10^{-1}$ $\bar{x}\bar{z} = 0.927 * (-0.235)10^5 = -0.2178 \cdot 10^5 \to v := \overline{\bar{x}\bar{z}} = -0.218 \cdot 10^5$
$\bar{x} + \bar{y} = (0.927 + 0.0000232)10^6 = 0.9270232 \cdot 10^6 \to \overline{(}\bar{x} + \bar{y}) = 0.927 \cdot 10^6$
$2 * \overline{(}\bar{x} + \bar{y}) = 1.854 \cdot 10^6 \to w := \overline{(}2 * \overline{(}\bar{x} + \bar{y}) = 0.185 \cdot 10^7$
$v - w = -(0.00218 + 0.185)10^7 = -0.18718 \cdot 10^7 \to \overline{(}v - w) = -0.187 \cdot 10^7$ Obtemos posto $r = xz - 2(x+y)$ que em $FP(10,3,-9,9)$: $\bar{r} = \overline{(}xz - 2(x+y)) = -0.187 \cdot 10^7$.
O resultado exato obtido com 10 digitos decimais é $r = -1.8765994944 \cdot 10^6 = -0.18765994944 \cdot 10^7$
Erro absoluto: $|r - \bar{r}| = |-0.18765994944 + 0.187|10^7 = 6.5994944 \cdot 10^3$
Erro relativo: $\frac{|r-\bar{r}|}{|\bar{r}|} = \frac{6.5994944 \cdot 10^3}{0.187 \cdot 10^7} \approx 0.003529 = 0.3529\%$

(2) Considere o problema de computar a seguinte operação matricial $A*B+C$, onde A, B, C são matrizes reais de dimensão 2 (ou seja, $A, B, C \in \mathbb{R}^{2\times 2}$)

a) Escreva um algoritmo (no papel) para computar $A * B + C$ para matrizes genéricas reais de dimensão 2. Sucessivamente escreva um código em uma linguagem de sua escolha que implemente o algoritmo apresentado.

b) Considere as matrizes de dimensão 2 a seguir:

$$\tilde{A} = \begin{pmatrix} 1.0001 & 0 \\ 1.9995 & 3 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} 4.002 & 0 \\ 0 & -1.3333 \end{pmatrix}, \quad \text{e} \quad \tilde{C} = \begin{pmatrix} 4 & 2 \\ -8 & 0.1 \end{pmatrix}.$$

– O seu algoritmo é estável para computar $\tilde{A} * \tilde{B} + \tilde{C}$ em $FP(10,3,-9,9)$?

– Compare o cálculo de $\tilde{A} * \tilde{B} + \tilde{C}$ em $FP(10, 3, -9, 9)$ usando arredondamento com respeito ao valor (exato) de $\tilde{A} * \tilde{B} + \tilde{C}$ (obtido, por exemplo, com o seu código executado em um computador).
– Pode dizer quando o algoritmo será estável ao variar dos coeficientes das matrizes? Dê um exemplo numérico que comprove a sua resposta.

Resoluçao (2):

a) **Require:** $A = (a_{ij})_{i,j=1,2}$, $B = (b_{ij})_{i,j=1,2}$, $C = (c_{ij})_{i,j=1,2}$ os coefficientes das matrizes

1: $d_{ij} \leftarrow \sum_{k=1}^{n} a_{ik} b_{kj}$ para cada $i, j = 1, 2$

2: $r_{ij} \leftarrow d_{ij} + c_{ij}$ para cada $i, j = 1, 2$

3: **return** $R = (r_{ij})_{i,j=1,2}$

Código em Matlab:

```
function R=atividade1ex2(A,B,C)
  D=A*B;
  R=D+C;
end

>>A= [1.0001  0;  1.9995  3];
>>B= [4.002  0;  0  -1.3333];
>>C= [4  2;  -8  0.1];
>>atividade1ex2(A,B,C)
```

b) – O algoritmo resulta ser instavel se $A * B$ tem elementos proximos de C, porque em tal caso pode aparecer o erro do cancelamento subtrativo que leva a um erro relativo alto respeito ao valor exato.

– O resultado exato (obtido com o código em cima) é

$$D = \tilde{A} * \tilde{B} + \tilde{C} = \begin{pmatrix} 4.0024002 & 0 \\ 8.001999 & -3.9999 \end{pmatrix} + \tilde{C} = \begin{pmatrix} 8.0024002 & 2.0 \\ 0.001999 & -3.8999 \end{pmatrix}$$

Em $FP(10, 3, -9, 9)$ obtemos $\tilde{A} = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}$, $\tilde{B} = \begin{pmatrix} 4 & 0 \\ 0 & -1.33 \end{pmatrix}$, $\tilde{C} = \begin{pmatrix} 4 & 2 \\ -8 & 0.1 \end{pmatrix}$,

e obtemos que $\overline{\tilde{A} * \tilde{B}} = \tilde{A} * \tilde{B} = \begin{pmatrix} 4 & 0 \\ 8 & -3.99 \end{pmatrix}$, e então $\overline{\tilde{A} * \tilde{B}} + C = \begin{pmatrix} 8 & 2 \\ 0 & -3.89 \end{pmatrix}$.

Portanto

$$\bar{D} := \overline{\tilde{A} * \tilde{B} + \tilde{C}} = \begin{pmatrix} 8 & 2 \\ 0 & -3.89 \end{pmatrix}$$

Notamos que o erro relativo no calculo dos elementos $(2, 1)$ esplode até o infinito: $E_r(d_{21}) = \frac{|0.001999 - 0|}{0} = +\infty$. Em vez nas outras componentes, onde não temos um cancelamento dos valores o erro resulta ser pequeno:

$E_r(d_{11}) = \frac{|8.0024002-8|}{8} = \frac{0.0024002}{8} = 3.00025 \cdot 10^{-4} \approx 0.03\%$,
$E_r(d_{12}) = \frac{|2-2|}{|2|} = 0, \quad E_r(d_{22}) = \frac{|-3.8999+3.89|}{|3.89|} \approx 0.00254 = 0.254\%$

– Para poder obter erros pequeno temos e assim um algoritmo estavel podemos escolher $\tilde{C}$ bastante diferente de $-\tilde{A}\tilde{B}$ por exemplo $\tilde{C} = -\tilde{A}\tilde{B} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

Assim com $\tilde{C} = \begin{pmatrix} -3 & 1 \\ -7 & 4.99 \end{pmatrix}$ obteremos

$$\bar{D} := \overline{\tilde{A} * \tilde{B} + \tilde{C}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

que tem até um erro relativo nulo em cada elemento da matriz, porque
$D := \tilde{A} * \tilde{B} + \tilde{C} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.